

(19) World Intellectual Property Organization  
International Bureau



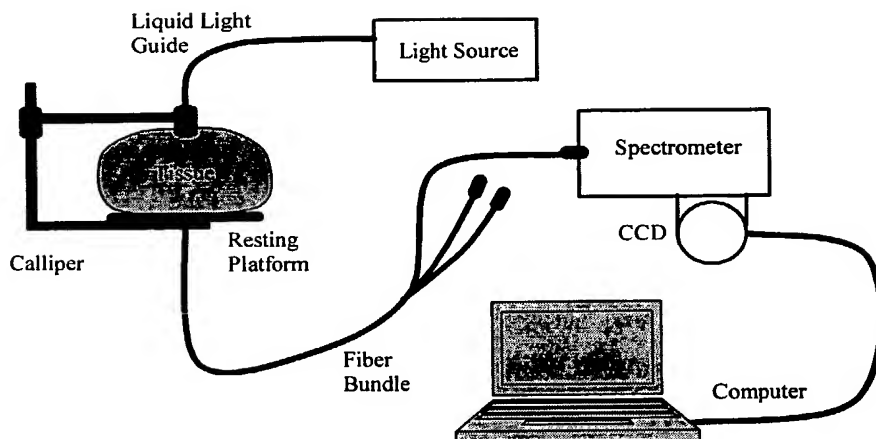
(43) International Publication Date  
30 May 2003 (30.05.2003)

PCT

(10) International Publication Number  
**WO 03/043492 A1**

- (51) International Patent Classification<sup>7</sup>: **A61B 5/00**
- (21) International Application Number: PCT/CA02/01771
- (22) International Filing Date:  
20 November 2002 (20.11.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/331,633 20 November 2001 (20.11.2001) US
- (71) Applicant (for all designated States except US): **UNIVERSITY HEALTH NETWORK** [CA/CA]; 610 University Avenue, Toronto, Ontario M5G 2M9 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **LILGE, Lothar** [DE/CA]; 913D Adelaide Street West, Toronto, Ontario M6G 3T2 (CA). **WILSON, Brian, C.** [GB/CA]; 85 Indian Grove, Toronto, Ontario M6R 2Y6 (CA). **SIMICK, Michelle, K.** [CA/CA]; 1550 Eagleview Drive, Pickering, Ontario L1V 5H6 (CA). **BOYD, Norman, F.** [CA/CA]; 162 Rosedale Heights Drive, Toronto, Ontario M4T 1C8 (CA).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report
  - before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: OPTICAL TRANSILLUMINATION AND REFLECTANCE SPECTROSCOPY TO QUANTIFY DISEASE RISK



(57) Abstract: The present invention uses spectroscopic tissue volume measurements using non-ionizing radiation to detect pre-disease transformations in the tissue, which increase the risk for this disease in mammals. The method illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range, detecting light transmitted through, or reflected from, volume of selected tissue, and obtaining a spectrum of the detected light. The spectrum of detected light is then represented by one or more basis spectral

components, an error term, and an associated scalar coefficient for each of the basis spectral components. The associated scalar coefficient is calculated by minimizing the error term. The associated scalar coefficient of each of the basis spectral components is correlated with a pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease to obtain the susceptibility for the mammal to developing the pre-selected disease.

WO 03/043492 A1

## **OPTICAL TRANSILLUMINATION AND REFLECTANCE SPECTROSCOPY TO QUANTIFY DISEASE RISK**

### **CROSS REFERENCE TO RELATED UNITED STATES PATENT APPLICATION**

This patent application relates to United States provisional patent application, Serial No. 60/331,633, filed on November 20, 2001, entitled OPTICAL TRANSILLUMINATION SPECTROSCOPY to quantify Disease Risk.

### **FIELD OF THE INVENTION**

The present invention relates to the use of optical transillumination/reflectance spectroscopy as a diagnostic for determination of the risk for diseases.

### **BACKGROUND OF THE INVENTION**

As early cancer or disease detection and diagnosis make further inroads into clinical practice, the high cost of current disease screening techniques redirected the focus of the investigators towards methods capable of quantifying the risk towards these diseases as a practical pre-screening tool. One particular application area is epidemiological science and public health research into cancer prevention, so the invention is applicable for a wide variety of chronic or slowly developing diseases such as Alzheimer's, or Multiple Sclerosis. However, prevention required identification of the population at risk in association with an appropriate prevention or risk reduction intervention for example through the control of disease inducing agents or lifestyle changes (exercise, diet etc.) An example risk quantification is the work related to assessing the breast cancer risk in the general population or subgroups thereof by Boyd, Yaffe et al. (1-3) which showed that X-ray density patterns are identified as having one of the highest odds ratios towards the risk of breast cancer between the low risk and the high risk groups. For the specific case of breast cancer, radiologically dense breast tissue on mammography indicates the presence of stromal and epithelial tissue in the breast, the amount of which is strongly related to risk of breast

cancer, with increasing amount of radiologically dense tissue related to increased risk. The ability to assess the breast cancer risk enables new steps in cancer prevention, for example through lifestyle and dietary changes (4).

One of the major disadvantages of the current standard for Breast Cancer Risk assessment is the use of ionizing radiation. This results in the late introduction of this diagnostic modality during the life of a woman, due to the inherent risk when using ionizing radiation in a diagnostic modality. Other good risk predictors are in general also only available once a woman has reached around 40 years of age, such as cancer incidence in first degree relatives (mother and sisters). However, the late onset in using these risk assessment modality will reduce the available time frame for any intervention aimed at reducing the disease risk. Hence, there is a clear need for a non-ionizing modality which can be employed in young patients, here in post puberty pre-menopausal women.

Non-ionizing radiation was employed in various optical mammography approaches, usually in attempts to image the breast, and to detect breast lesions (5,6) commonly using frequency domain technologies at only a few specialized wavelengths, or as spectroscopic approaches for the determining the tissue optical properties of normal versus malignant breast tissue (7). These spectroscopic applications, including an article by Egan and Dolen (8) are rather intended for determination of the probability for the presence of cancer, but do not address the concept of risk assessment, e.g. as a pre-screening tool.

United States Patent No. 6,121,775 is directed to an MRI imaging method and apparatus and provides a physical interrogation methods related to detecting small changes in tissue.

United States Patent No. 5,079,698 is directed to a transillumination method and apparatus for the diagnosis of breast tumors and other breast lesions by normalization of an electronic image of the breast. United States Patent No. 6,002,958 is directed to a method and apparatus for diagnostics of internal organs. This patent teaches the use of NIR radiation in the 0.6-1.5  $\mu\text{m}$  wavelength range and adds ultrasound to the analysis tools. These two patents specifically create images of the organ

United States Patent No. 6,095,982 discloses a spectroscopic method and apparatus for optically detecting abnormal mammalian epithelial tissue' covers only Raman and fluorescence methods. United States Patent No. 6,069,689 discloses an apparatus and methods relating to optical systems for diagnosis of skin diseases while very generally written and addressing, reflectance, fluorescence and Raman, using a plurality of light emitting diodes. While some changes in the tissue (skin) are mentioned the idea of risk assessment is not included in this or any other patent related to the use of non-ionizing radiation.

### SUMMARY OF THE INVENTION

The present invention uses spectroscopic tissue volume measurements using non-ionizing radiation to detect pre-disease transformations in the tissue, which increase the individuals risk for this disease.

In one aspect of the present invention there is provided a method for assessing susceptibility for developing a pre-selected disease in a mammal, comprising:

- a) illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range;
- b) detecting light transmitted through, or reflected from, said volume of selected tissue, and obtaining a spectrum of said detected light;
- c) representing the spectrum of detected light by a set of basis spectral components, an error term, and an associated scalar coefficient for each basis spectral component in said set, the set of basis spectral components including at least one basis spectral component, the associated scalar coefficient for each basis spectral component being calculated by minimizing the error term; and
- d) correlating the associated scalar coefficient for each spectral component with a pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease to obtain the susceptibility for the mammal to developing the pre-selected disease.

In another aspect of the invention there is provided an apparatus for assessing susceptibility for developing a pre-selected disease in a mammal, comprising:

a) holder means for holding and immobilizing an anatomical part of a mammal containing tissue to be optically probed;

b) light source means for illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range;

c) detection means for detecting light transmitted through, or reflected from, said volume of selected tissue;

d) computer control means connected to said detection means for producing a spectrum of said detected light from an output of said detection mean, the computer control means including processing means for representing the spectrum of detected light by a set of basis spectral components, an error term, and an associated scalar coefficient for each spectral component in said set, the set of basis spectral components including at least one basis spectral component, the associated scalar coefficient for each basis spectral component being calculated by minimizing the error term, the processing means includes means for correlating the associated scalar coefficient for each basis spectral component with a pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease to obtain the susceptibility for the mammal to developing the pre-selected disease, the computer control means including display means for displaying the susceptibility.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The method and apparatus constructed in accordance with the present invention will now be described, by way of example only, reference being had to the accompanying drawings, in which:

Figure 1 shows the block diagram of one possible embodiment of an apparatus produced in accordance with the present invention;

Figure 2 show a reconstruction of an experimentally obtained optical transillumination spectrum by a linear combination of either only the first two or four principle components.

Figure 3 shows an example of a typical set of measurements comprised of eight spectra from a volunteer representing the four quadrant of the breast on either side of the bilateral organ;

Figure 4 shows the correlation of the  $t_1$  and  $t_2$  scores from the spectra that were repeatedly measured resulting in a regression slope and correlation coefficient close to unity;

Figure 5 showing the first 4 components obtained from the PCA following data pre-processing which included thickness and transfer function correction, note that component 2 show inverse absorption for the lipid and water peaks respectively;

Figure 6 shows the resulting weights or loading factors  $t_1$  vs.  $t_2$  for the first 2 principle components for thickness and transfer function correction as pre-processing option, square and rhombus symbols represent high and low breast density subjects, respectively;

Figure 7 shows the first four components obtained from the PCA following data pre-processing which includes only transfer function correction;

Figure 8 shows the resulting weights or loading factors  $t_1$  vs.  $t_2$  for the first two principle components using transfer function correction as only pre-processing option, square and rhombus symbols represent high and low breast density subjects, respectively;

Figure 9 shows the first 4 components obtained from the PCA following data pre-processing which includes thickness and transfer function correction followed by autoscaling of the data;

Figure 10 shows the resulting weights or loading factors  $t_1$  vs.  $t_2$  for the first 2 principle components using thickness and transfer function correction followed by autoscaling of the data, square and rhombus symbols represent high and low breast density subjects, respectively;

Figure 11 assigns physical meaning to the four quadrants in the  $t_1$  vs.  $t_2$  plots, shown here for thickness and transfer function corrected and autoscaled data

Figure 12 shows an identification of spectra form low and high density breast-tissue within a 3D plot of  $t_1$  vs.  $t_2$  and  $t_3$  using thickness and transfer function correction as pre-processing options;

Figure 13 shows the predicted % parenchymal tissue density based on PLS analysis of the optical transillumination spectra versus percent density according computer assisted analysis of the mammograms, solid symbols refer to the training set and open symbols to the validation set;

5        Figure 14 shows spectrally resolved transillumination and reflectance measurements are obtained by the used of one of more appropriate VIS/NIR light sources and opto-electronic detector systems measuring the surface proximal volume layer and the total volume separately; and

10        Figure 15 shows an example of a three wavelength frequency domain system to determine the differential path-length factor of photons traversing tissue.

### DETAILED DESCRIPTION OF THE INVENTION

Referring to Figure 1 shows an apparatus 10 used in the present  
15        invention comprises a light source 12, waveband selection devices, delivery optical fibre 14, tissue support 16, collection optical fibre bundle 18, wavelength dispersing element, opto-electronic detector 20 and CPU 22. Spectra can be corrected for the wavelength dependent transfer function of the instrument and the physical interoptode distance. It will be understood that  
20        this is not required as it can also be convoluted into the numerical correlation analysis. Additionally, the optical fibre 14, the tissue support and the optical fibre bundle is optional. The wavelength selective device can be either prior to or post the interrogated tissue.

For use in studies related to breast tissue, the light source 12 was a 12  
25        Watt halogen lamp (Welch Allyn) with stabilized power supply serving as a broadband light source. The ultraviolet to green and mid infrared radiation were eliminated using a cut-off filter (at 500 nm) and a heat filter, respectively. The remaining light was coupled into a 5mm diameter liquid light guide, which served as the source fiber in contact with the skin on the top of the breast. A  
30        total power of 0.25 Watt, covering the 550 - 1300 nm bandwidth, was delivered to the skin surface. The holder of the source fiber and the plate in which the detector fiber was embedded were made of black plastic to absorb reemitted photons . The source fiber was in firm contact with the tissue and slightly compressed it by 5mm. The source and detector fibers (optodes) were

held coaxially, pointing towards each other, via a custom-made calliper attached to the resting platform. The calliper allowed the measurement of the inter-optode distance (cm) and hence, the physical thickness of the interrogated tissue. Transmitted light was collected via a custom-made 7 mm diameter fiber bundle (P&P Optical Kitchener, Ontario, Canada). Wavelength resolved detection in the visible-near infrared was achieved using a holographic transillumination grating (15.7 rules/mm blazed at 850 nm) (Kaiser, California, USA) and a 2D, cryogenically-cooled silicon CCD (Photometrics, New Jersey, USA) at a spectral resolution of less than 3 nm between 625 nm and 1060 nm. This spectral resolution was achieved by positioning a 0.5 mm slit between the distal end of the collection fiber and the spectral grating. The peak quantum efficiency ( $>0.8$ ) of the detector was at 780 nm, falling to 0.2 at 1100 nm. By imaging the entrance slit of the spectrograph onto the 2D CCD, 50 rows of pixels were effectively exposed to detected light, thus increasing the dynamic range of the electronic detection by up to 50 times. To minimized background noise, cryogenic cooling reduces the dark counts to  $\sim 0.06$  electrons per hour. Further noise reduction was achieved using exposure times of 2-3 seconds and averaging up to five scans. The system's dynamic range was  $>5$  OD (optical densities) with a signal-to-noise ratio of  $>10^4$  at peak sensitivity and  $>10$  across the remaining bandwidth of the system.

In order to collect the transmitted light emerging from the tissue, a wavelength dispersing medium (in the preferred embodiment a grating) for the various wavelength ranges use, if present, and an opto-electronic detector, here a Si based ( $0.65\ \mu\text{m}$  to  $1.5\ \mu\text{m}$ ) photodiode array (PDAs) are used. The PDA's information is transferred directly or delayed to the computer containing a numerical analysis program to quantify the disease risk using a previously trained algorithm. The correlation between spectral data and disease risk is based on the spectral attenuation, through photon absorption and scattering, only. The particular disease risk assessment system comprises in its current version a continuous optical spectrum with wavelength in the 550 nm to 1000 nm range.

Different embodiments of the apparatus may include an InGaAs based ( $0.9\ \mu\text{m}$  to  $1.7\ \mu\text{m}$ ), or equivalent opto-electronic detector to cover also the

longer wavelength range which is transmitted through tissue, or can comprise also only various wavelengths or wavelength bands within the VIS/NIR range, such as attainable with filters of direct emission devices with limited bandwidth such as laser, LED or similar.

5           In another embodiment of the apparatus there is included at least a single  
(whereas three is preferred) wavelength, frequency domain transillumination system, spanning most of the steady state wavelength range for transillumination. In one non-limiting embodiment the three wavelengths may  
10   be 785, 808, 905 nm. The lasers in the frequency domain system need to be modulated between 10 to 400 MHz (215MHz in the present system) to obtain a good phase shift resolution for 5 cm physical pathlength between delivery and detection optode. The system should provide a phase shift resolution of better than  $2^\circ$ . The lasers used here provide ~200 mW average power at the  
15   distal end of the delivery fibers in contact with the patient. Intensity modulation is achieved by modulating the driving current of the lasers via a biased-T, a method to recapture the transmitted light and guide it to the PMT with fast time response, a preamplifier filtered to the appropriate modulation frequency, a lock-in detection to quantify the demodulation and the phase shift compared  
20   to a standard either from the laser driver current or a portion of the light prior to entering the tissue, and a data transfer system to a computer to translate the phase shift and demodulation data into a differential path length factor and scattering coefficient. In frequency domain measurements various mod  
operandi can be employed one being heterodyne detection technique, which  
25   can use also very lower power lasers. Other methods to quantify the optical pathlength are described (10, 11) and include time resolved detection or time resolved single photon detection.

30           The ability to quantify disease risk and hence the population members at risk can be increased through additional measurements such as skin reflectance measurements to be able to subtract the skin contribution to the optical transillumination spectrum, so clearly this information may not be required. Depending on the part of the body being diagnosed additional components such as anatomical supports may be included, for example if the breasts are being studied, a support stand for the breast, and a holder for the

liquid light guide and optical fibres for light delivery, and a random fibre bundle for light collection. The latter may be a trifurcated fibre bundle to deliver transmitted photons to the PMT (frequency domain measurements), Si-based and InGaAs-based PDAs respectively. Again these are to a large extent optional, for example one can envision a cup like device which holds the sources and detectors in direct contact with the tissue. A means to measure the angle and distance between the delivery and detection optodes equivalent to the physical tissue thickness or in the case of a reflectance measurement the interoptode distance as the latter will determine the tissue depth most likely interrogated by the photons. It is noted that the inventors have shown some correlation and predictive abilities with thickness correction of the data. Software to extract spectral absorption and light scattering features to quantify the risk for a specific disease is loaded onto the CPU. Software was trained using data from clinical studies providing the optical spectra and an independent measure of the risk as discussed hereinafter.

In order to correct for variations in the wavelength dependent signal transfer function of the system, all transillumination spectra were referenced to a daily collected transmission standard made of 1 cm thick, ultra-high density polyurethane (Gigahertz Optics, Munich, Germany). All measured spectra are thus given as wavelength dependent (rel OD). This referencing to a known standard yields a universal applicable dataset and hence the subsequent developed mathematical models correlating transillumination spectra with risk will also become universally valid, that is independent of the actual instrument used to collect the data.

The spectral volume measurements can be augmented by adding extended long wavelength (comprising the NIR range transmitted through tissue, e.g. up to 1.7  $\mu\text{m}$ ) and frequency domain measurements. The former providing more information about vibrational bands of the biomolecules in the interrogated volume. The latter enables by using a limited number of wavelengths (3 or more), to determine the differential pathlength factors of the tissue (6). Combining this with the transillumination spectra allows one to obtain absolute absorption spectra and subsequently derivation of the contributions of various chromophores responsible for the absorption spectra

measured providing additional information for the identification of the population at risk and also insight into the molecular changes associated with or resulting from the tissue transformation.

## **ANALYSIS OF REFLECTANCE OR TRANSILLUMINATION OPTICAL DATA FOR ASSESSMENT OF DISEASE RISK**

The following describes four nonlimiting examples of mathematical modelling techniques capable of establishing a correlation between the optical spectra and a particular outcome, in the present case risk of disease. This is not a complete or comprehensive list of all the methods available which may be used and other methods such as hybrid linear analysis (12) are available and persons skilled in this type of analysis can identify other methods. These spectral analysis techniques have been used extensively in chemometrics field to solve for the concentration of a constituent of interest without knowing the spectra of all constituents present in a chemical sample (13) or where multiple chemicals of interest have overlapping spectra, (14). These analysis techniques construct models that identify the variance within the spectral data set or, when trained with a known range of concentrations, can identify the variance that is relevant to the constituent(s) of interest. Through training the model derives component spectra that resemble the constituent spectra within the sample, and can later be used to predict the concentrations of constituents from new sample spectra. However, these component spectra may not have direct physical meaning but can represent the spectral features of the pure constituent spectra.

In the case of breast cancer which is exemplified hereinafter, it has been shown by Thompson and Tatman (9) that the tissue composition will change slowly as the breast is undergoing sequential changes towards dysplasia, carcinoma *in situ* and then invasive cancer, often involving or resulting in concentration change of different optically (light absorption or fluorescence) active molecules and/or structures (light scattering).

In the specific case of breast cancer the idea is based on the fact that the same structures and chemicals, that give rise to the x-ray attenuation, the current (gold) standard for breast cancer risk assessment, will also result in

changes in the transillumination spectra. These changes will be evident in the attenuation and/or scattering of the visible/NIR light due to different contributions of the optically active molecules. Correlation between the x-ray mammography and the spectral transillumination information can be established and quantified by a variety of different numerical methods, among which are principal component analysis or linear discriminate analysis when the breast density based on mammographic analysis is provided as classification (nominal data) or using partial least squares or principle component regression when the mammographic analysis is provided as % dense tissue (interval data). Besides x-ray mammographic analysis, other methods which may be used to obtain the parenchymal breast tissue density may include ultrasound, computed tomography, cone beam computed tomography, electrical impedance spectroscopy and magnetic resonance imaging.

#### **Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)**

PCA is optimized for comparison of vectors with nominal data, in this case correlating spectra to the density classification, thereby making use of the complete spectral information while reducing the dataset size. For this, PCA determines the amount of variance within the test group of spectra and uses it to iteratively reduce the dataset to representative spectra, called components. The spectral data are represented by a smaller number of vectors in a lower dimensional space while still including the maximum amount of variance within the data set. (15). This is accomplished by solving for the covariance or correlation matrix of the data matrix  $\mathbf{X}$  ( $m \times n$ ) comprised of the dataset of all spectra obtained ( $n$ ) and the spectral range monitored ( $m$ ), such that:

$$\text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{n-1} \quad (1)$$

PCA decomposes the data matrix  $\mathbf{X}$  as the sum of the outer products of the vectors  $\mathbf{t}_i$  and  $\mathbf{p}_i$  and a residual matrix  $\mathbf{E}$ :

$$\begin{aligned} \mathbf{X} &= \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \mathbf{t}_3 \mathbf{p}_3^T + \dots + \mathbf{t}_I \mathbf{p}_I^T + \mathbf{E} \\ \text{or } \mathbf{X} &= \mathbf{T} \mathbf{P}^T + \mathbf{E} \end{aligned} \quad (2)$$

Where the elements of the  $\mathbf{t}_i$  ( $n \times 1$ ) vectors are the scores that contain information on how the spectra relate to each other, and the  $\mathbf{p}_i$  vectors ( $m \times 1$ ) or components are the eigenvectors of the covariance matrix and show how the selected variance relates to each other. The components  $\mathbf{p}_i$  are eigenvectors of the covariance matrix, so that

$$\text{cov}(\mathbf{X}) \mathbf{p}_i = \lambda_i \mathbf{p}_i \quad (3)$$

where  $\lambda_i$  is the eigenvalue associated with the eigenvector  $\mathbf{p}_i$ . Also for any  $\mathbf{X}$  and  $\mathbf{t}_i$  and  $\mathbf{p}_i$  pair  $\mathbf{X} \mathbf{p}_i = \mathbf{t}_i$  (i.e. the score vector  $\mathbf{t}_i$  is the linear combination or representation of the original  $\mathbf{X}$  defined by  $\mathbf{p}_i$ ). It is generally found that data can be described in fewer components  $\mathbf{p}_i$  than original variables ( $m$ ) and that the components can be combinations of variables that are useful descriptions of particular constituents in the tissue. As stated above, the shape of the useful components will be a combination of spectral signatures of chromophores that vary with tissue density. Scores (elements of  $\mathbf{t}_i$ ) that differentiate between tissue densities identify useful components ( $\mathbf{p}_i$ ).

The scores (elements of  $\mathbf{t}_i$ ) can be graphically plotted against one another to show any potential clustering of spectra that are related. In this study PCA was calculated on the test set (2/3 of available spectra) and the same mathematical model, e.g. retaining the  $\mathbf{p}_i$  was used to predict the scores  $\mathbf{t}_i$  on the validation set (1/3 spectra).

Linear Discriminant Analysis (LDA) was used on the PCA scores to enhance the differentiation between the two extreme nominal categories (low vs high density). LDA finds a discriminant rule for defined groups within the dataset. It is based on drawing the boundary halfway between the means on a pair wise basis (adjusted slightly if there is some prior probability of group identity). It calculates the discriminant rule by computing optimizing factors (*criterion<sub>j</sub>*) that are based on the covariance matrix of the given groups. The targets that are used for training identify the density category group (16)

$$criterion_j = \text{inv}(\text{cov } X_j) \cdot S_b \quad (4)$$

Where  $j$  is the particular group and  $S_b$  is the between class variance defined by

$$S_b = \sum_j (x_j - \bar{x}_n)^2 \quad (5)$$

5 In this study, the LDA algorithm was trained using the training dataset by leaving one spectrum out during each training run and then predicting it's tissue density group. The rest of the spectra from the validation set were then classified using the trained model. The LDA scores can also be plotted in a similar cluster diagram as the PCA scores (16) Table 1 shows the key  
10 differences between PCA and LDA.

Statistical significance for the PCA and LDA prediction was established using measures similar to sensitivity and specificity values commonly used to evaluate the validity of diagnostic tests. As transillumination is currently envisioned as a method for identifying those women with high tissue density  
15 within the entire population, increased HDM is preferable over increased LDM. Consequently, a high density measure (HDM) was used to assess the quality of the prediction model. HDM was defined as the ratio of those women that were predicted to have high density tissue with the optical  
20 transillumination method to those who were categorized as having high tissue density by the Radiologist. Conversely, the low density measure (LDM) is a measure of the ability to correctly identify those spectra that do not have high density.

25 **Table 1:** Properties of Principal Component Analysis and Linear Discriminant Analysis.

	PCA	LDA
Input data	Spectral vector	Requires PCA scores as input
Result space size e.g. classifications	Undefined since # of PC's is undefined	Defined by training procedures

Data reduction	Vector->scalar	Scalar->scalar
Training or Density readings needed	No	Yes
Presentation of output data	Relationship of scalars by cluster analysis	Relationship of scalars by categories or cluster analysis
What variance is used	Those captured in principal components	The variance between the defined groups
Additional output information	Component vectors contain physical meaning	Discriminant rule no apparent physical meaning

Figure 2 shows a reconstruction of an experimentally obtained spectrum by either only two or four principle components.

#### **Principal Component Regression and Partial Least Squares Analysis**

5        PCR and PLS are analysis techniques are optimized for comparison of vector inputs with interval target data. For this case the spectra are compared to the percent density readings. Statistical significance was established using the Pearson correlation coefficient as well as the slope and intercept of the regression line. Both PCR and PLS analysis require training and perform regression analysis on the dataset, but use different types of variance identification to predict the target variable, in this study the percent density read from the mammogram, see Table 2.

10        As a first attempt for solving for a concentration of a constituent, Inverse Least Squares analysis has been used for well behaved data, where spectral features are not overlapped and predictions of the constituent is not effected by collinearity problems with other constituents in the sample. Here **Y** is defined as:

$$\mathbf{Y} = \mathbf{X} \mathbf{b} \quad (6)$$

Where **b** is the regression vector that is found by solving for the pseudoinverse of **X**. This cannot be used here since all the spectral features contributing to the spectral shape and that lead to density prediction are not defined so there could be collinearity problems.

5 Principle Component Regression calculates principal components in a similar manner to Principal Component Analysis and then performs regression analysis of the scores matrix with the known targets or tissue densities from the training set. It does this by solving the following equation

$$\mathbf{Y} = (\mathbf{P} \mathbf{T}^T) \mathbf{b} \quad (7)$$

$$10 \quad \text{where } \mathbf{b} = \mathbf{P} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y} \quad (8)$$

which is similar for Inverse Least Squares analysis above in equation 7, except that  $\mathbf{X} = \mathbf{P} \mathbf{T}^T$ . The regression vector **b** is the least squares solution for **T**, where **T** are scores from the PCA and **T** is a  $(n \times k)$  matrix, where *n* is the number of spectra and *k* is the number of factors used in the model. **Y** is the  
15  $(n \times 1)$  vector of targets, in this case percent density of the region of interest of the mammogram (17).

For Partial Least Squares Analysis, the basis vectors or latent variables (equivalent to the principal components) are extracted using the related targets, such as percent density, by solving equation 7 and relating **b** to the  
20 targets directly by solving:

$$\mathbf{b} = \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y} \quad (9)$$

where **W** is a weighting vector that relates the target set (prediction values) with the variance in the dataset during the decomposition into factors. This allows the PLS algorithm to solve for components that are correlated with the  
25 targets (**Y**) while describing a large amount of the variation in **X**. (17) For a more detailed explanation please see Kowalski BR and Geladi P (1986),(17) Wise BM (2000),(15) and Haaland DM and Thomas EV (1988). (14).

With both PCR and PLS only a single regression vector **b** is used in subsequent predictions of spectra. This **b** vector is selected by calculating the  
30 predicted residual error sum of squares (PRESS) value for each **b<sub>i</sub>** using the training set. The chosen **b** vector has the fewest number of iterations (lowest value of *i*) and the lowest PRESS value when the training set is used to establish the model.

5

**Table 2:** Properties of Principal Component Regression and Partial Least Squares Analysis.

	PCR	PLS
Inputs	Density (targets) and spectra for a training set	Density (targets) and spectra for a training set
Data reduction	No	No
Density readings needed	For each spectrum in training set	For each spectrum in training set
Presentation of output data	Estimated percent density	Estimated percent density
What variance is used	Those captured in principal components	The variance that matches the targets
Additional output info	b vector	b vector

10

15

In this study the PRESS values were calculated for both PCR and PLS for the training sets, using the built-in cross-validation function in Matlab 6.0 PLS toolbox. The **b** vector with the lowest PRESS value was selected. The training set was randomized and split into four sections, the respective algorithms were then trained by a Venetian blind method with four repeats. The subsequent **b** vectors were then averaged and used to predict the tissue density for entire training set and subsequently the validation set.

The method disclosed herein will now be described with a nonlimiting example of study in breast tissue.

20

## EXAMPLE

### BREAST TISSUE TRANSILLUMINATION STUDY

Transillumination spectroscopy measurements were taken in a dark room with the volunteer seated comfortably, and the breast resting on a horizontal support (Figure 1). Total data acquisition time was approximately 15 minutes allowing measurements at all 8 positions and necessary re-positioning of the optodes. There was little discomfort to the volunteers due to point compression of the breast by the source optode.

Figure 3 shows an example of a typical set of measurements comprised of eight spectra from the volunteer. Spectra from the respective bilateral positions of the breast show symmetry, as expected in women with healthy breast tissue. In previous studies, this criterion of spectral symmetry in the same position of both breasts was used, by Egan and Dolan, to indicate absence of breast cancer in the examined areas (8). The reproducibility of the optical transillumination measurements within one visit was analyzed by recruiting a volunteer to undergo multiple procedures.

A total of 8 spectra were collected per volunteer representing the four quadrants (medial, distal, lateral and central) of both breasts. These spectra can be further pre-processed by correcting for tissue thickness (OD/cm) and/or auto-scaling prior to development and testing of the mathematical models used to then correlate with tissue density. In general for Principal Component Analysis and similar models, scaling is used if the input variables have different units or if there are large deviations in some input variables compared to other variables within the data set.

For auto-scaling, the data was mean-centered, where the mean optical density value of the data set at each wavelength was subtracted from all spectra, and scaled to one unit variance by dividing all the values at each wavelength by the standard deviation for that wavelength. This was applied only to the training set data for PCA models developed, whereas scaling of the validation set the previous obtained mean and standard deviation vectors were used. See Table 3 for pre-processing options used for the interpretation of the spectral data set.

**Table 3.** Pre-processing techniques of the spectral dataset and a reference number used in future sections.

Pre-processing of Dataset	
Transfer function corrected	I
Thickness and transfer function corrected	II
Autoscaled and transfer function corrected	III
Autoscaled and thickness and transfer function corrected	IV

5

A detailed example of establishing a correlation between optical spectra and disease risk follows for the example of PCA. It is generally found that data can be described in fewer components  $p_i$  than original variables ( $m=436$  wavelength elements) and that the components can be combinations of variables that are useful descriptions of particular constituents in the tissue. As stated above, the shape of the useful components will be a combination of spectral signatures of chromophores that vary with tissue density. Scores (elements of  $t_i$ ) that differentiate between tissue densities identify useful components ( $p_i$ ). The scores (elements of  $t_i$ ) can be graphically plotted against one another to show any potential clustering of spectra that are related. Here the PCA algorithm was trained on a test set ( $n=544$  spectra) and the same mathematical model, i.e. retaining the  $p_i$ , was used to predict the scores  $t_i$  on the validation set ( $n=192$  spectra).

10

15

20

Statistical significance for the PCA prediction was established using high-density measure (HDM) as it is preferable over increased low density measure (LDM) both are similar to sensitivity and specificity. As transillumination is envisioned as a method for identifying women with high tissue density within the entire population, improved HDM is desired.

Consequently, HDM was used to assess the quality of the prediction model defined as the ratio of women predicted to have high density tissue to those who were categorized as having high tissue density by the radiologist. LDM is a measure of the ability to correctly identify those spectra that do not have high density.

Figure 4 shows the correlation of the  $t_1$  and  $t_2$  scores from the spectra that were repeatedly measured. Figure 4 also shows acceptable reproducibility with slopes of the regression and the  $R^2$  values close to one. There is a spread in the component scores ( $t_1$  and  $t_2$ ), but there is a clustering of the position related data.

Figure 5 shows the principal components ( $p_i$ ) from the PCA using  $n = 544$  spectra (thickness and transfer function corrected) obtained to date. These first four components contain 97.6 %, 1.2 %, 0.6 % and 0.3 % of the variance in the total data set, respectively, for a combined total of 99.8% of the variance. The cluster plot of the scores for  $t_1$  and  $t_2$  is shown in Figure 6. This plot illustrates discrimination of the breast tissue areas across a diagonal line in the  $t_1$  vs.  $t_2$  space.

Non-thickness corrected spectra were used to determine the effect of thickness on the robustness of the PCA model prediction. The component spectra for  $p_1$ - $p_4$  are very similar to the thickness-corrected component plot; see Figures 5 and 7 respectively. The  $t_1$  vs.  $t_2$  plot, Figure 8, for this data shows discrimination as a function of  $t_2$  only. Similarly component spectra can be seen in Figure 9 for thickness and transfer function corrected data that were additionally autoscaled, with the  $t_1$  vs.  $t_2$  cluster plot shown in Figure 10. The scores for both  $t_1$  and  $t_2$  are centered on zero as expected but show similar clusters as Figure 8.

HDM and LDM were determined to compare which spectral pre-processing option best differentiates the high and low density spectra in cluster plots according to Figure 6, 8 and 10. HDM and LDM data shown in the Table 4 were calculated from the same training and validation data sets separately.

**Table 4:** HDM and LDM of Principal Component Analysis results for test and validation set measurements.

Pre-processing of Dataset		Test set		Validation Set	
		HDM	LDM	HDM	LDM
Transfer function					
corrected (Figure 8)	I	85 %	97.0 %	88 %	90 %
Thickness and transfer function					
corrected (Figure 6)	II	88 %	93 %	93 %	89 %
Autoscaled – transfer function					
corrected (data not shown)	III	86 %	94 %	90 %	86 %
Autoscaled – thickness and transfer function					
corrected (Figure 10)	IV	87 %	92 %	93 %	90 %

Autoscaling removes some spectral information since the mean spectrum is wavelength dependent. As the spectral features contributing to the discrimination between high and low breast density or risk are unknown, losing spectral information is very unwise. This is reflected in a reduced HDM and LDM as seen in Table 4 for pre-processing option III and IV.

The scores  $t_1$  and  $t_2$  resulting from thickness corrected and non - corrected component spectra demonstrated that it is possible to differentiate between the subjects having low or high breast tissue density subjects (Figures 8 and 10).

While mathematical models derived for thickness or non-thickness corrected optical spectra differentiate high and low breast tissue densities, their  $t_1$  vs.  $t_2$

cluster plots differ. One explanation for the difference in the thickness corrected vs. non-corrected cluster plots is the effect of the physical tissue thickness on the overall variance within the spectral data set and therefore the determination of the components by the PCA algorithm. Thickness of the tissue contributes non-uniformly to individual spectra as larger breasts tend to contain more fatty tissue.

In the model of the data that was not thickness corrected,  $p_1$  did not differentiate between high and low density tissue. This can also be seen in the magnitude of the  $t_1$  in Figure 5 versus the non-thickness corrected component spectra in Figure 7. This indicates that the thickness values contribute to the magnitude of  $p_1$  masking other contributions that do differentiate between tissue densities, such as light scattering. This leaves  $t_2$  as the only component to preserve information to distinguish the density of the breast.

When comparing the autoscaled versus non-autoscaled data, there were minimal changes in the principal component spectra and minor differences in HDM and LDM values, see Figures 6 and 10 and Table 4. Autoscaling as part of the pre-processing can degrade regions with flat or extreme spectral variation.(14) In this case, degraded spectral features could include regions of the spectrum with minimal wavelength dependence and hence a first derivative close to zero. For example, the haemoglobin inflection points are more pronounced in the non-autoscaled data than in the autoscaled components. Conversely, the large spectral features of water and lipids are large compared to other structures in the spectra, but are less pronounced after autoscaling. In this study, the only differences in the HDM values are in the training set with the non-autoscaled data having almost 2% higher scores for both HDM and LDM.

Principal components can reveal particular regions of the spectrum that represent important physical properties or entities within the tissue that contribute to differentiation. Component spectra 1 ( $p_1$ ) and 2 ( $p_2$ ) are the most important cover the highest amount of variance in the data set. While components 3 and 4 have similar or inverse shape as component 2 they take less variance into account.

Principal component 1 ( $p_1$ ) has a small dependence on wavelength and has negative values in each prepared model. The spectrally flat

characteristics resulting from thickness corrected spectra can be attributed to attenuation due to absorption, scattering, and, therefore optical pathlength and losses at the tissue boundary. The surprisingly flat spectral shape of the scattering contribution is due to the derivation of OD used here, based on the wavelength dependent transfer function calibration by a polyurethane block which exhibits also high Mie scattering, e.g. Mie scattering is present in the spectrum and hence cancels itself out. We propose that  $p_1$  carries optical pathlength information despite not showing the typical  $\lambda^{-1}$  dependency, (18) but contains information to determine breast density through the overall scattering power. Low density tissue spectra have a reduced amount of scattering compared to high density tissue, and, therefore, higher values of  $t_1$  in Figures 8 and 10. This relationship in scattering properties is also seen in the scattering coefficient data by Peters *et al.* (19) and Troy *et al.* (20) supporting this interpretation of the first principal component.

The other component enabling differentiation between low and high tissue densities  $p_2$ , has a more complex shape when compared to  $p_1$ . The most important spectral features in the component are the lipid with inverse water peaks present at 930 nm and 980 nm, respectively. When  $t_2$  values are positive, the lipid peak is the dominant spectral feature as anticipated for fatty or low density tissue. Spectra from the high density tissue have negative  $t_2$  values and water absorption becomes the dominant structure in the component spectrum. Graham *et al.* (21) also observed this relationship between water and density values when using MRI to quantify percent density. In their study the water content of the tissue was measured directly and showed good correlation to percent tissue density ( $r = 0.79$ ). (21)

Spectral feature contributions by haemoglobin can be seen between 625 and 850 nm within  $p_2$  where the negative slope and inflection points of the haemoglobin curve are apparent. Dense breast tissue has large negative scores ( $t_2$ ) compared to the low density tissue. Haemoglobin contributes to the variance in  $p_2$ , with elevated contribution of blood related absorption for high water absorption seen in the higher density tissue. This simultaneous appearance of water and haemoglobin absorption can be explained physiologically, as tissues with higher water content and hence cellular content, require improved vascular supply and, thereby, increased blood

volume. (22) Since positive  $t_2$  scores are related to low tissue density and positive  $t_1$  scores are related to low tissue scatter, the cluster plot of  $t_1$  and  $t_2$  can be divided into quadrants as shown in Figure 11, highlighting the relationship between the spectral features and the known physical attributes of the breast tissue.

Even though p3 and p4 did not show differentiation between high and low density tissue by themselves in a 2-dimensional plot like p1 and p2, areas of the component spectra can be interpreted. Additionally p3 shows a lipid absorption peak, but water and hemoglobin absorption are almost absent additionally the lipid peak is shifted towards longer wavelength, so the reason therefore is unknown. p4 shows influence from the hemoglobin, with the same slope but inverse inflection points to p2. Even though differences in the amplitude and general shape of the curves are minimal when compared to p2 the magnitude of the scores  $t_3$  and  $t_4$  are much smaller than that of the first two components. One cannot exclude specifically p4 as a third contributor to increase HDM and LDM when using a three dimensional analysis as suggested in figure 12.

Figure 13 shows an example of representing the parenchymal tissue density on an interval scale (% dense area in the mammogram) rather than a nominal scale (low, medium, high density). The % dense area can be determined using a computer assisted imaging program by a trained observer such as for example a radiologist. In the case presented here no trained person was utilized and hence the percent density from the mammograms contained a large error in the repeat measure which ultimately limits the accuracy of the presented correlation. However, it serves to indicate that the variance reduction programs employed are capable of providing a single integer number to the physician monitoring his patients' health.

The results disclosed herein show that *In vivo* optical transillumination spectroscopy is a technically feasible method and capable of predicting breast tissue densities with acceptable correlation to mammographic densities as an indirect method of cancer risk assessment. Thus, transillumination spectroscopy may offer a novel "first step" in the risk assessment of healthy women regardless of menstrual cycle, age, ethnic background or menopausal status as the data here was not stratified by either event.

HDM and LDM values close to or above 90 % are encouraging to distinguish between low and high density tissues. These HDM and LDM values are higher compared with other physical examinations, such as ultrasound (23) and magnetic resonance imaging, (21) reported to be between 70-80 %.

Optical transillumination spectroscopy offers the potential of a real-time and cost-effective method compared with ultrasound and has the ability to quantify a large range of tissue densities for breasts that are up to 6 cm in thickness. An added advantage of transillumination spectroscopy over ultrasound and MRI is the fact that results are derived from preset mathematical models and, hence, no additional highly trained personnel are required for assessment. This reduces the overall cost to the healthcare system for this risk-assessment technique. A painless procedure and the inherent safety of this method will likely contribute to a high compliance rate.

X-ray mammography uses ionizing radiation and is considered unacceptable as a tool to assess breast density for women less than forty years of age and for frequent measurement. However, transillumination spectroscopy is safe for women of all ages. This allows risk assessment to commence at a much younger age when the life style and diet are perhaps easier to influence have more time to exert their beneficial effects to reduce the risk and could ultimately lead to reduced incidence rates.

While optical transillumination spectroscopy is a promising tool to monitor the effectiveness of chemopreventive, dietary or lifestyle studies for the reduction of breast cancer risk, its ability to detect physical changes over a period of time in the breast tissue of a given individual needs to be demonstrated in a prospective longitudinal study.

The predictive value of optical spectroscopy for disease susceptibility quantification can be increased through additional measurements which can include extending the optical waveband, obtain optical information of interrogated tissue not contributing to the disease risk and obtaining information separately for light scattering.

The first option includes the use of wavelength up to 1.7  $\mu\text{m}$  which contain among others additional water and lipid absorption bands using an

InGaAs based opto-electronic detector or an equivalent system. This is particular of interest for large tissue volumes. If the optically interrogated tissue volume is small, that is in the range of several  $\text{mm}^3$ , the short wavelength band using light of approximately 360 nm to 600 nm may provide the relevant information.

For the second option, using a combined transmittance and reflectance measurement as shown in Figure 14 enables collection of the optical information related to the superficial tissue, here skin, separately through the reflectance technique, whereas the transillumination spectroscopy contain information about the superficial and deep tissue.

As described above frequency or time domain reflectance and transillumination measurements can provide an effective optical path length through phase shift or lifetime measurements. The optical path length relative to the physical path length between the optodes is related to the light scattering power of the tissue. Figure 15 shows an example of an embodiment of a system comprised of three diode lasers to measure the breast tissue light scattering power. The remaining part of the spectrum can be interpolated as shown in the paper by Cerussi et al (24).

The present method and apparatus disclosed herein has been exemplified using breast cancer as the disease of interest which involves correlating the associated scalar coefficient of the basis spectral component(s) with the pre-selected property of parenchymal breast tissue density known to be indicative of susceptibility of breast tissue for breast cancer.

It will however be understood that this invention is not restricted to use in assessing risk for breast cancer but many other diseases as well. For example, the method disclosed herein is contemplated to be applicable for correlating optical information of other mammalian tissue with risk factors associated with diseases such as neurodegenerative diseases including Multiple Sclerosis, Alzheimer's and Parkinson's diseases; oncology including, prostate, rectal and testicular cancers, autoimmune diseases including Sinustisus, rheumatoid arthritis, and Chron's disease. In each case the relevant tissues are optically sampled and the basis spectral component(s) are obtained using other techniques such as ultrasound, X-ray analysis,

magnetic resonance imaging, potential molecular markers indicating initial changes in the tissue, or epidemiologically derived questionnaires proven to correlate with the disease of interest as just a couple of examples. The present invention is applicable to mammals in general and is not restricted to humans.

As used herein, the terms "comprises", "comprising", "includes" and "including" are to be construed as being inclusive and open ended, and not exclusive. Specifically, when used in this specification including claims, the terms "comprises", "comprising", "includes" and "including" and variations thereof mean the specified features, steps or components are included. These terms are not to be interpreted to exclude the presence of other features, steps or components.

The foregoing description of the preferred embodiments of the invention has been presented to illustrate the principles of the invention and not to limit the invention to the particular embodiment illustrated. It is intended that the scope of the invention be defined by all of the embodiments encompassed within the following claims and their equivalents.

#### References:

1. Boyd NF. Lockwood GA. Martin LJ. Knight JA. Jong RA. Fishell E. Byng JW. Yaffe MJ. Tritchler DL. Mammographic densities and risk of breast cancer among subjects with a family history of this disease. *Journal of the National Cancer Institute*. 91(16):1404-8, 1999
2. Boyd NF. Lockwood GA. Byng JW. Tritchler DL. Yaffe MJ. Mammographic densities and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*. 7(12):1133-44, 1998
3. Byng JW. Yaffe MJ. Jong RA. Shumak RS. Lockwood GA. Tritchler DL. Boyd NF. Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics*. 18(6):1587-98, 1998
4. Knight JA. Martin LJ. Greenberg CV. Lockwood GA. Byng JW. Yaffe MJ. Tritchler DL. Boyd NF. Macronutrient intake and change in mammographic density at menopause: results from a randomized trial. *Cancer Epidemiology, Biomarkers & Prevention*. 8(2):123-8, 1999

5. Zhu Q. Conant E. Chance B. Optical imaging as an adjunct to sonograph in differentiating benign from malignant breast lesions  
Journal of Biomedical Optics. 5(2):229-36, 2000
6. Franceschini MA. Moesta KT. Fantini S. Gaida G. Gratton E. Jess H.  
5 Mantulin WW. Seeber M. Schlag PM. Kaschke M. Frequency-domain techniques enhance optical mammography: initial clinical results.  
Proceedings of the National Academy of Sciences of the United States of America. 94(12):6468-73, 1997
7. Quaresima V. Matcher SJ. Ferrari M. Identification and quantification of  
10 intrinsic optical contrast for near-infrared mammography.  
Photochemistry & Photobiology. 67(1):4-14, 1998
8. Egan and Dolen Optical spectroscopy. Pre-mammography marker.  
Acta Radiologica. 29(5): 497-503, 1988
9. Thompson and Tatman Physiological and pathological factors of  
15 human breast disease that can influence optical diagnosis. Ann New York Acad Scie 838:171-93 1998
10. Gratton E. Fantini S. Franceschini MA. Gratton G. Fabiani M.  
Measurements of scattering and absorption changes in muscle and brain.  
20 Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences. 352:727-35, 1997
11. Jacques SL. Path integral description of light transport in tissue.  
Annals of the New York Academy of Sciences. 838:1-13, 1998
12. Berger, A.J., Koo, T-W., Itzkan, I., & Feld, M.S. 1998, 'An Enhanced  
25 Algorithm for Linear Multivariate Calibration' Analytical Chemistry, vol. 70(3), pp. 623-627.
13. Haaland DM and Thomas EV., Partial Least-Squares Methods for Spectral Analysis: 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information *Anal. Chem.* 60  
30 1193-1202, 1988
14. Haaland DM and Thomas EV., Partial Least-Squares Methods for Spectral Analysis. 2. Application to Simulated and Glass Spectral Data *Anal. Chem.* 60 1202-1208, 1988

15. Wise BM., PLS Toolbox Tutorial: Matlab Version 6. Eigenvector  
Research Inc. 2000
16. Balakrishnama S, Ganapathiraju A. Linear Discriminant Analysis: A  
Brief Tutorial Institute for Signal and Information Processing Mississippi  
State University 2001
17. Kowalski BR and Geladi P., Partial Least-Squares Regression: A  
Tutorial. *Analytica Chimica Acta* 185: 1-17, 1986
18. Welch AJ and van Gemert MJC (Editors) Optical-Thermal Response of  
Laser-Irradiated Tissue. Plenum Press New York, USA. 1995
19. Peters J, Patterson MS, Wilson BC, Optical properties of normal and  
diseased human breast tissues in the visible and near infrared. *Phys  
Med Biol* , 35:1317-1334, 1990
20. Troy T, Page DL, Sevick-Muraca EM., Optical Properties of Normal  
and Diseased Breast Tissues: Prognosis for Optical Mammography.  
*Biomedical Optical Spectroscopy and Diagnostics J Opt Soc Am*, 3: 59-  
66, 1996
21. Graham SJ, Bronskill MJ, Byng JW, Yaffe MJ, Boyd NF., Quantitative  
correlation of breast tissue parameters using magnetic resonance and  
X-ray mammography. *Br J Cancer*;73(2):162-8, 1996
22. Haskell RC, Svaasand LO, Tsay TT, Feng TC, McAdams MS,  
Tromberg BJ., Boundary conditions for the diffusion equation in  
radiative transfer. *J Opt Soc Am A* 11: 2727-41, 1994
23. Kaizer L, Fishell EK, Hunt JW, Foster FS, Boyd NF.  
Ultrasonographically defined parenchymal patterns of the breast:  
relationship to mammographic patterns and other risk factors for breast  
cancer. *Br J Radiol*. 61(722): 118-24, 1988
24. Cerussi AE, Berger AJ, Bevilacqua F, Shah N, Jakubowski D, Butler J,  
Holcombe RF, Tromberg BJ, Sources of Absorption and Scattering  
contrast for near-infrared optical mammography. *Acad Radiol* 8:211-  
218:2001

**THEREFORE WHAT IS CLAIMED IS:**

1. A method for assessing susceptibility for developing a pre-selected disease in a mammal, comprising:
  - a) illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range;
  - b) detecting light transmitted through, or reflected from, said volume of selected tissue, and obtaining a spectrum of said detected light;
  - c) representing the spectrum of detected light by a set of basis spectral components, an error term, and an associated scalar coefficient for each basis spectral component in said set, the set of basis spectral components including at least one basis spectral component, the associated scalar coefficient for each basis spectral component being calculated by minimizing the error term; and
  - d) correlating the associated scalar coefficient for each basis spectral component with a pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease to obtain the susceptibility for the mammal to developing the pre-selected disease.
2. The method according to claim 1 wherein the set of basis spectral components is obtained by the steps of
  - obtaining a range of values of the pre-selected property from a population cross section which is representative of the mammalian population at large, but who are not expressing the pre-selected disease;
  - in each mammalian member of said population cross section obtaining at least one spectrum by illuminating a volume of selected tissue with light having wavelengths covering the pre-selected spectral range, and detecting light transmitted through, or reflected from, said volume of selected tissue;
  - extracting common spectral elements from the at least one spectrum of each member of said population cross section and producing at least one set of component spectra from the extracted common spectral elements, the at least one set of component spectra including at least one component spectra;

identifying a set of component spectra which gives the highest predictive value for the pre-selected property of the selected tissue known to be indicative of susceptibility for the pre-selected disease; and

using said set of component spectra which gives the highest predictive value for the pre-selected property of the selected tissue known to be indicative of susceptibility for the pre-selected disease as the set of basis spectral components in step c) of claim 1.

3. The method according to claim 1 or 2 wherein the step of extracting common spectral elements from the at least one spectrum of each member of said population cross section and producing at least one set of component spectra from the extracted common spectral elements includes using one of Principle Component Analysis, Partial Least Squares and Principle Component Regression.

4. The method according to claim 1, 2 or 3 wherein the step of calculating the associated scalar coefficient by minimizing the error term includes using Chi-square fitting to minimize the error term.

5. The method according to claim 1, 2, 3 or 4 including repeating steps a) to d) inclusively for a particular mammal periodically over a pre-selected period of the mammal's lifetime and storing the susceptibilities calculated periodically to detect a rate of change in susceptibility for increased risk for the pre-selected disease.

6. The method according to claim 1, 2, 3, 4 or 5 wherein said mammal is a post puberty human female, and wherein the pre-selected disease is breast cancer, and the pre-selected property is parenchymal breast tissue density.

7. The method according to claim 6 wherein the parenchymal breast tissue density is measured by one of x-ray mammography, ultrasound, computed tomography, cone beam computed tomography, electrical impedance spectroscopy and magnetic resonance imaging.

8. The method according to claim 1 wherein the spectrum of said detected light is a continuous spectrum.
9. The method according to claim 1, 2, 3, 4, 5, 6, 7 or 8 wherein the spectrum of said detected light is a discrete spectrum.
10. The method according to claim 1, 2, 3, 4, 5, 6, 7 or 8 wherein the step of illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range includes illuminating with broad band light, discrete wavelengths or wavelength bands.
11. The method according to claim 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 wherein the steps of illuminating a volume of selected tissue and detecting light transmitted through, or reflected from, said volume of selected tissue includes adjusting an angle and a distance between a position on the tissue surface which is illuminated and a position on the tissue surface where the transmitted or reflected light emanates from which is detected.
12. The method according to claim 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 wherein said pre-selected disease is selected from the group consisting of neurodegenerative diseases, oncology based diseases and autoimmune diseases.
13. The method according to claim 12 wherein said neurodegenerative diseases include Multiple Sclerosis, Alzheimer's and Parkinson's disease, and wherein said oncology based diseases include prostate, rectal and testicular cancers, and wherein said autoimmune diseases include sinusitis, rheumatoid arthritis, and Chron's disease.
14. The method according to claim 2 wherein the set of basis spectral components includes two or more component spectra.
15. The method according to claim 1 wherein step d) includes using an effective mathematical variance reduction model to identify the basis spectral

components which through the associated scalar coefficient shows best correlation with the pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease.

16. An apparatus for optical assessing susceptibility for developing a pre-selected disease in a mammal, comprising:

a) holder means for holding and immobilizing an anatomical part of a mammal containing tissue to be optically probed;

b) light source means for illuminating a volume of selected tissue of a mammal with light having wavelengths covering a pre-selected spectral range;

c) detection means for detecting light transmitted through, or reflected from, said volume of selected tissue;

d) computer control means connected to said detection means for producing a spectrum of said detected light from an output of said detection mean, the computer control means including processing means for representing the spectrum of detected light by a set of basis spectral components, an error term, and an associated scalar coefficient for each spectral component in said set, the set of basis spectral components including at least one basis spectral component, the associated scalar coefficient for each basis spectral component being calculated by minimizing the error term, the processing means includes means for correlating the associated scalar coefficient for each basis spectral component with a pre-selected property of the selected tissue known to be indicative of susceptibility of the tissue for the pre-selected disease to obtain the susceptibility for the mammal to developing the pre-selected disease, the computer control means including display means for displaying the susceptibility.

17. The apparatus according to claim 16 wherein the at least one basis spectral component is obtained by the steps of

obtaining a range of values of the pre-selected property from a population cross section which is representative of the mammalian population at large, but who are not expressing the pre-selected disease;

in each mammalian member of said population cross section obtaining at least one spectrum by illuminating a volume of selected tissue with light having wavelengths covering the pre-selected spectral range, and detecting light transmitted through, or reflected from, said volume of selected tissue;

extracting common spectral elements from the at least one spectrum of each member of said population cross section and producing at least one set of component spectra from the extracted common spectral elements, the at least one set of component spectra including at least one component spectra;

identifying a set of component spectra which gives the highest predictive value for the pre-selected property of the selected tissue known to be indicative of susceptibility for the pre-selected disease; and

using said set of component spectra which gives the highest predictive value for the pre-selected property of the selected tissue known to be indicative of susceptibility for the pre-selected disease as the set of basis spectral components in claim 16.

18. The apparatus according to claim 16 wherein the light source means includes either one of a white light source such as a filtered halogen lamp, a series of bandwidth limited light emitting diodes (LEDs), several laser light sources all emitting within the pre-selected spectral range.

19. The apparatus according to claim 16 wherein the detection means includes either wavelength selective means in conjunction with a charge coupled device or a photodiode array all for parallel detection of the pre-selected wavelength band or an avalanche photodiode or point photo detector for sequential detection of the pre-selected wavelength band.

20. The apparatus according to claim 16, 17, 18 or 19 wherein the processing means includes calibration means for measuring a wavelength dependent instrument transfer function, and wherein said processing means includes means for removing instrumental affects on the spectrum of detected light.

21. The apparatus according to claim 19 wherein the calibration means for measuring a wavelength dependent instrument transfer function includes a pre-selected volume of material which preferentially scatters light rather than absorbing light, positioned in place of the pre-selected tissue.

22. The apparatus according to claim 21 wherein the material which preferentially scatters light rather than absorbing light is comprised of highly density poly-urethane.

23. The apparatus according to claim 16 wherein said detection means is positioned to detect light reflected from the tissue being illuminated, including means for adjusting the angle and distance between the light source means and the detection means.

24. The apparatus according to claim 16, 17, 18, 19, 20, 21, 22 or 23 wherein the light source means includes a light source and an optical fiber with a proximal end positioned to couple light into said optical fiber and a distal end from which light emanates from the fiber positioned so as to direct light onto the pre-selected tissue.

25. The apparatus according to claim 16, 17, 18, 19, 20, 21, 22, or 23 wherein the mammal is a post puberty human female, and wherein the pre-selected disease is breast cancer, and wherein the holder means includes a breast support including at least one illumination means and at least one detection means for each breast.

1/15

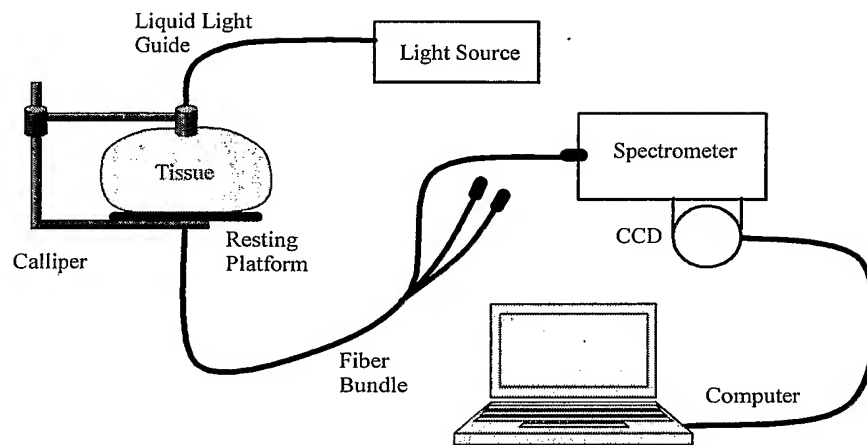


FIGURE 1

2/15

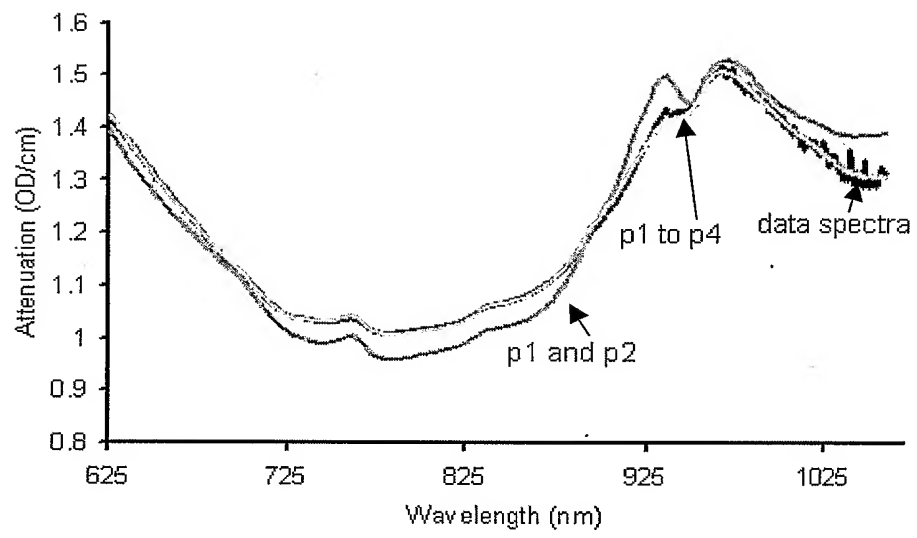


FIGURE 2

3/15

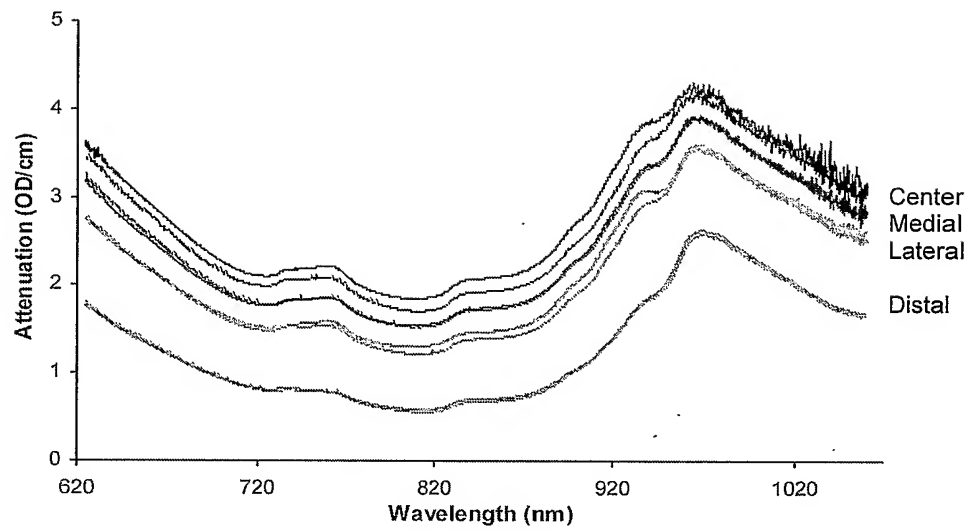


FIGURE 3

4/15

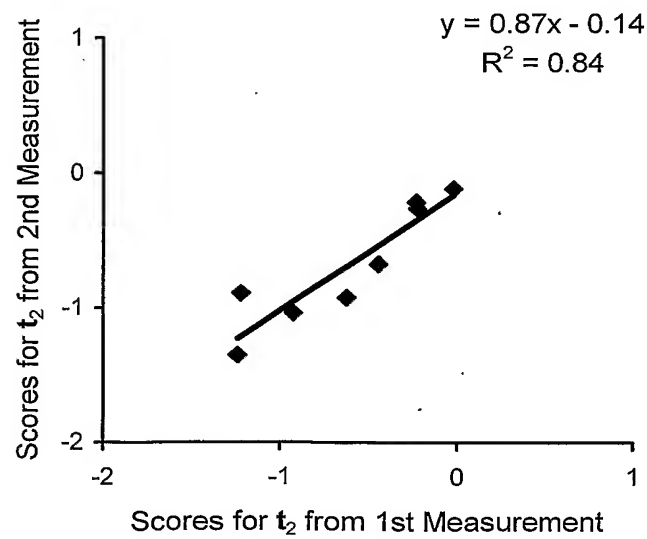
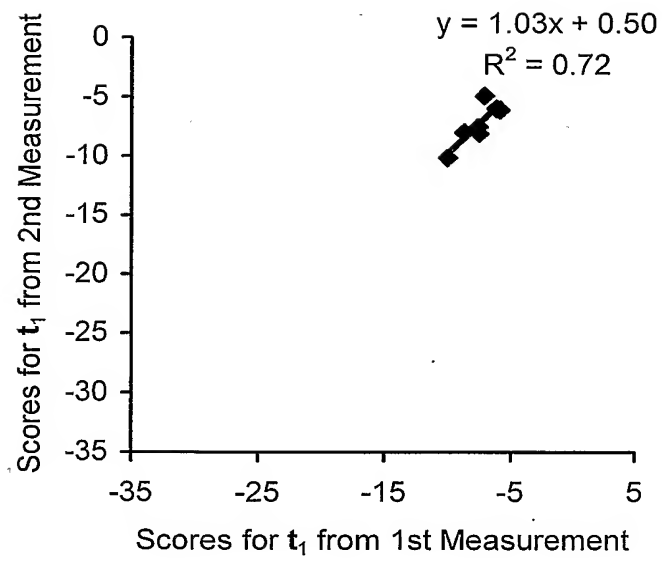


FIGURE 4

5/15

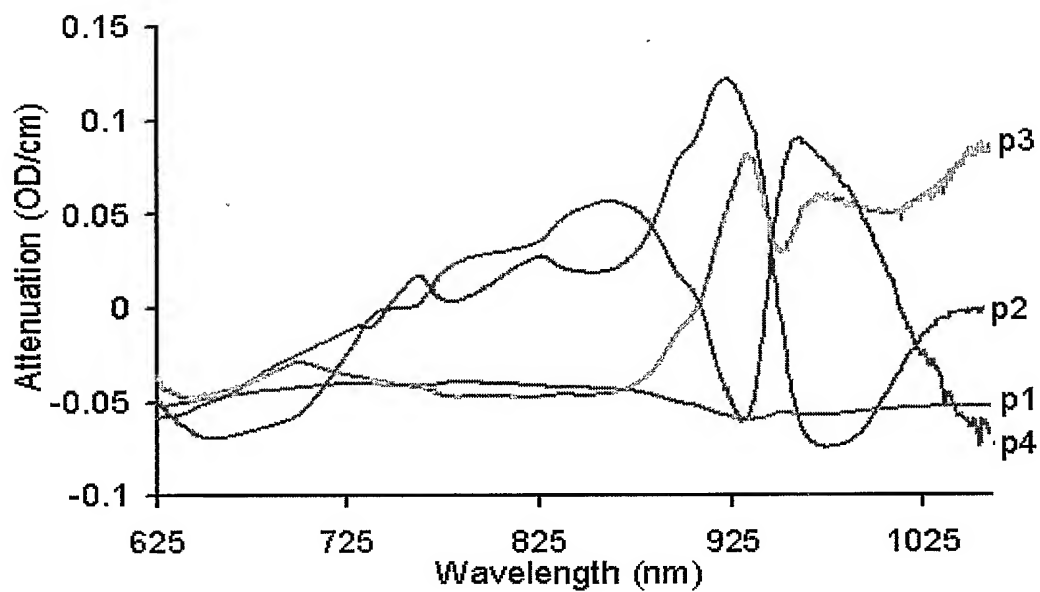


FIGURE 5

6/15

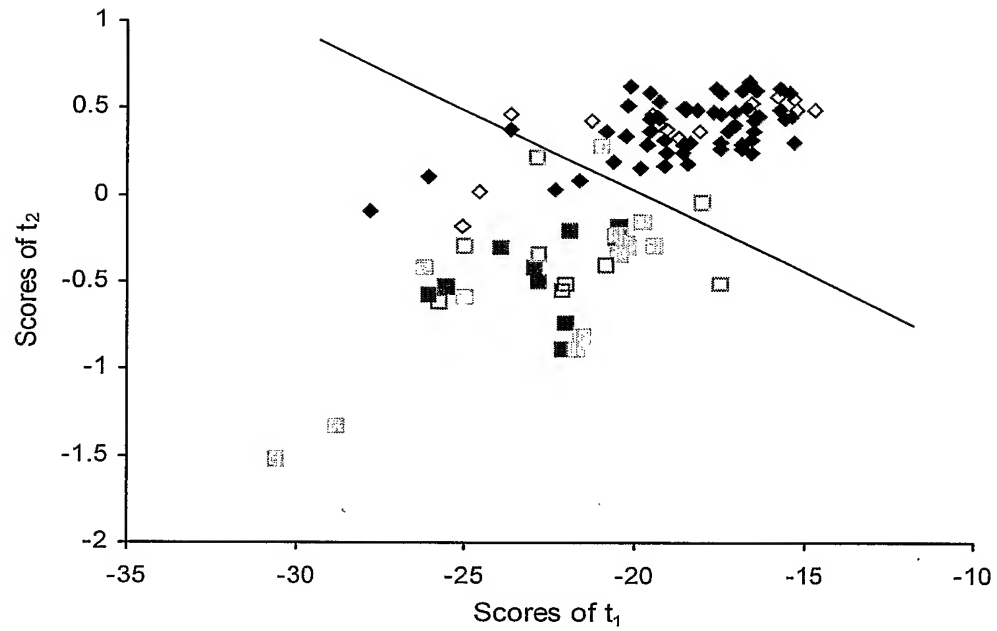


FIGURE 6

7/15

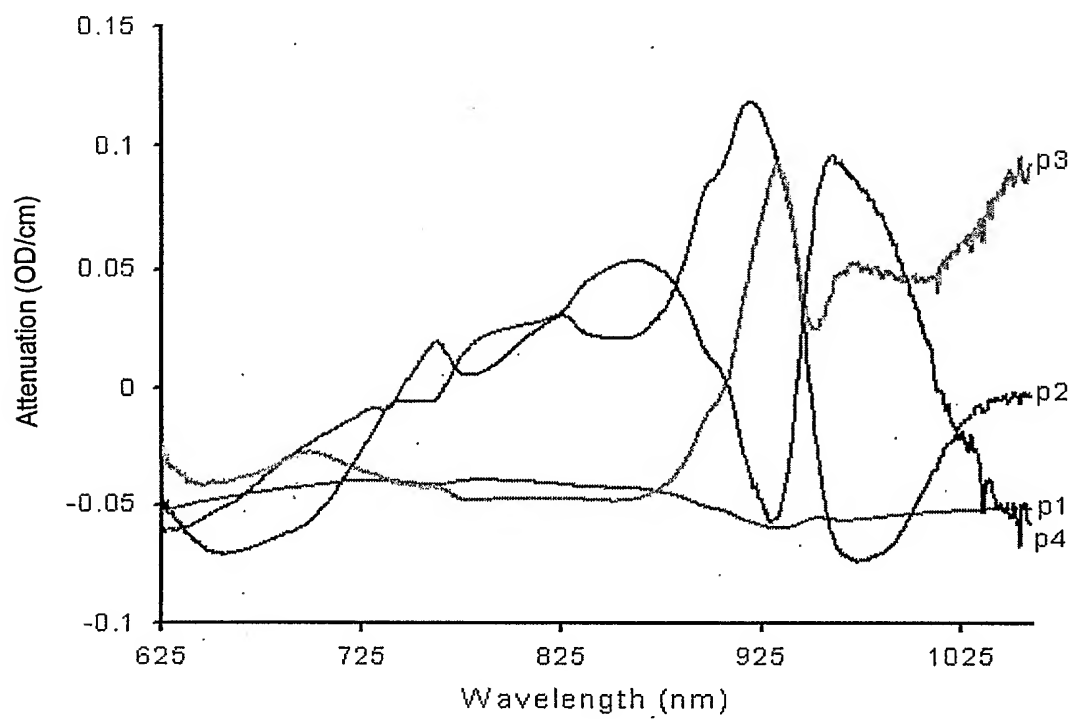


FIGURE 7

8/15

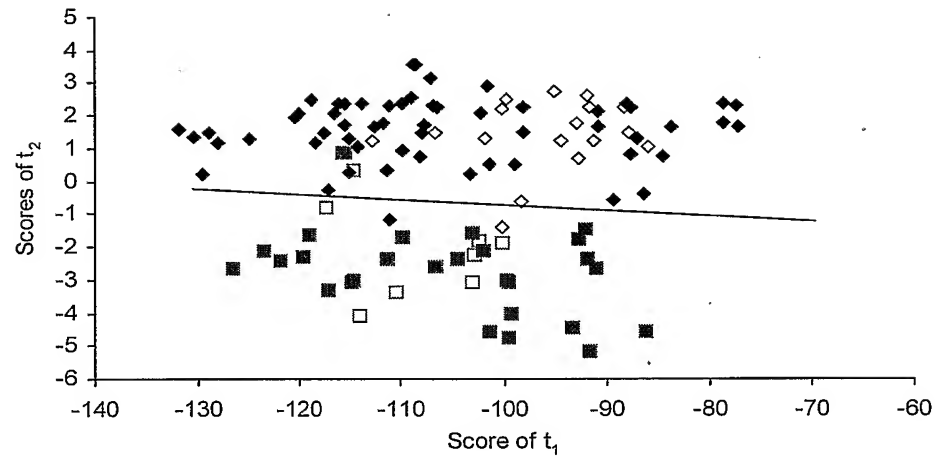


FIGURE 8

9/15

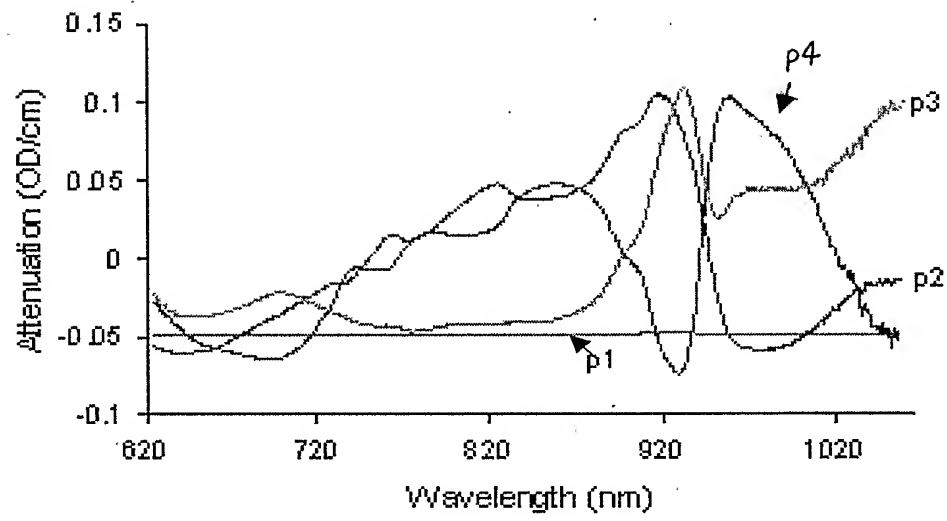


FIGURE 9

10/15

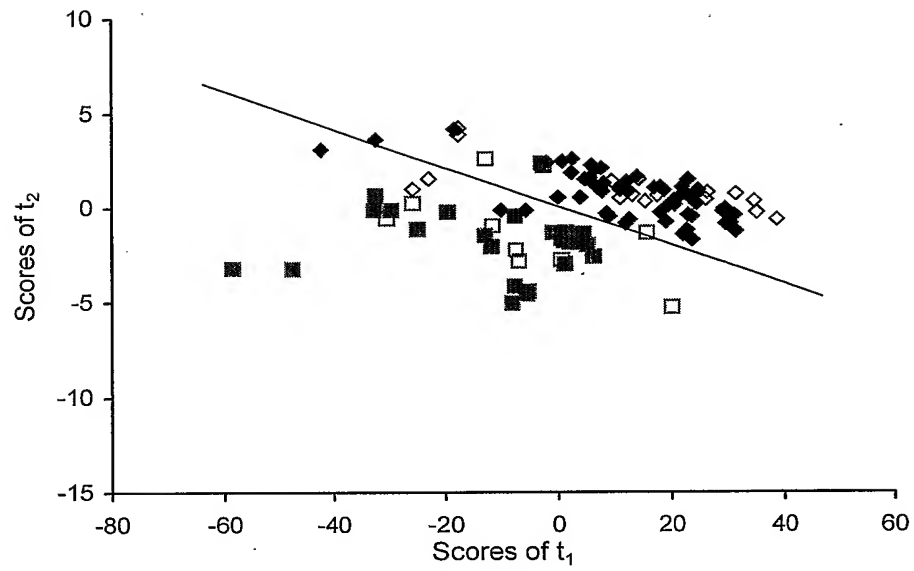


FIGURE 10

11/15

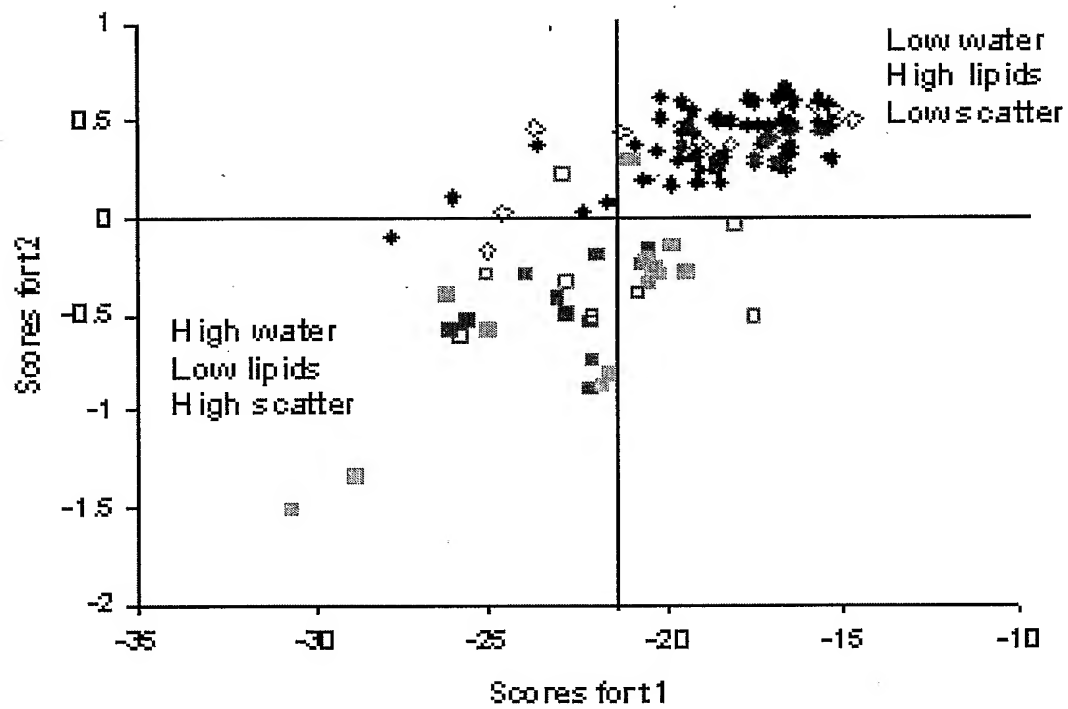


FIGURE 11

12/15

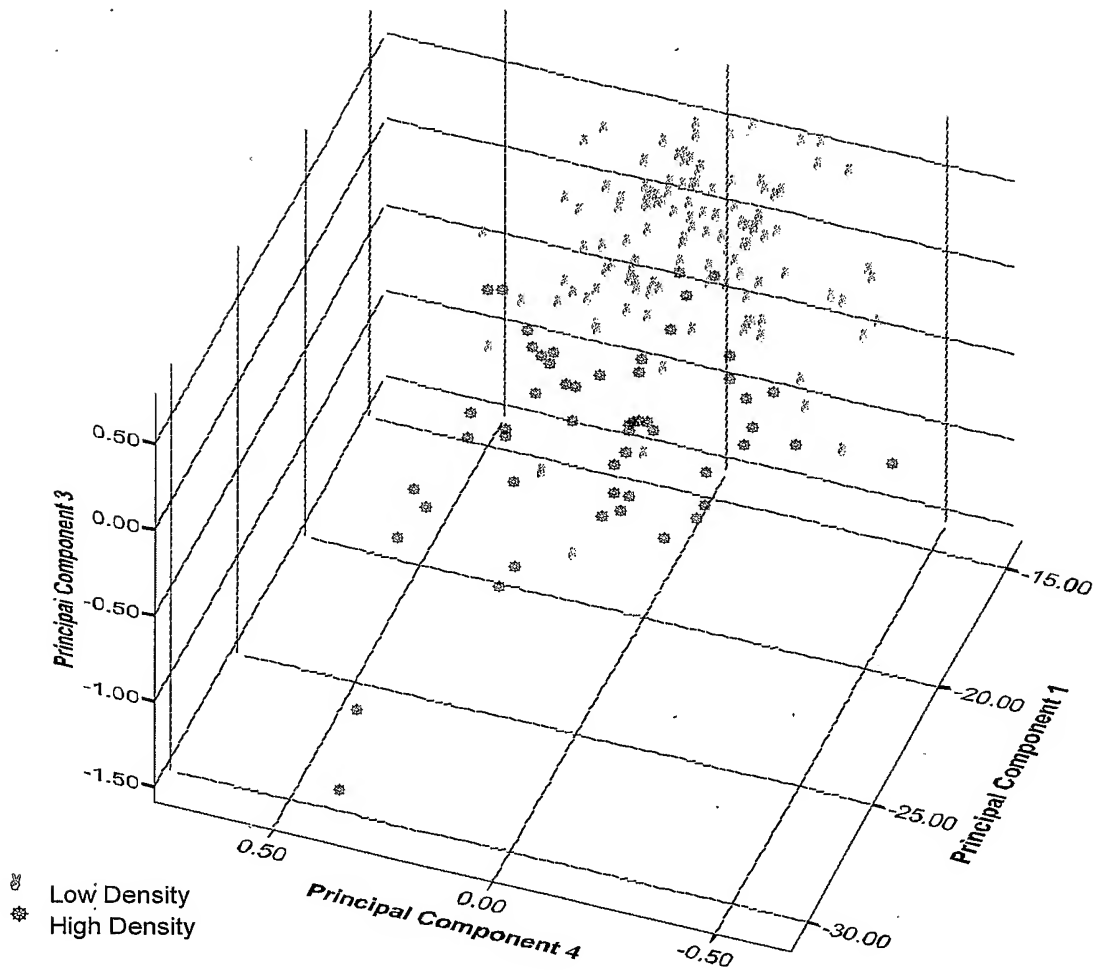


FIGURE 12

13/15

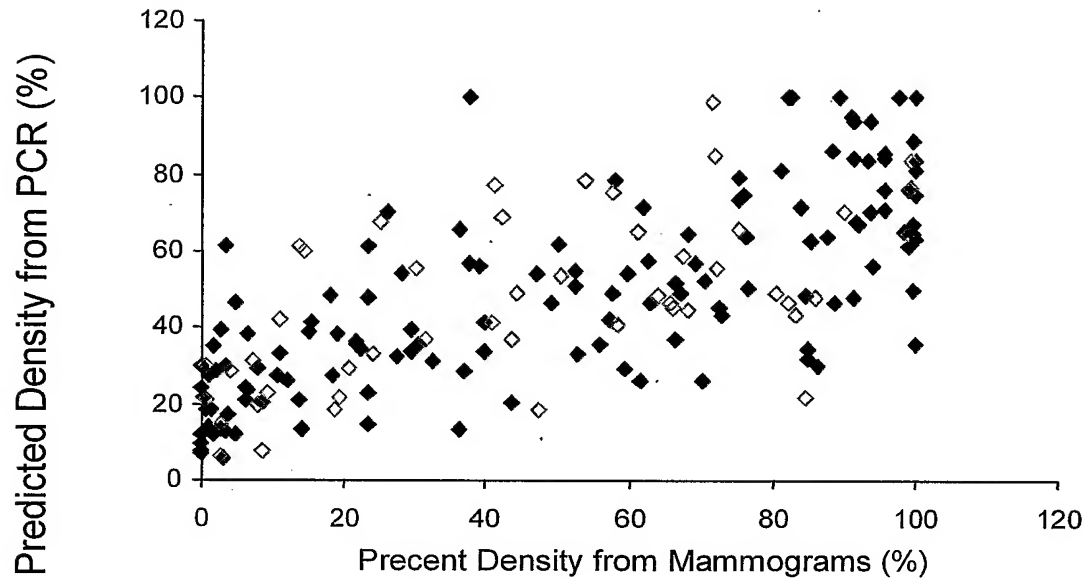


FIGURE 13

14/15

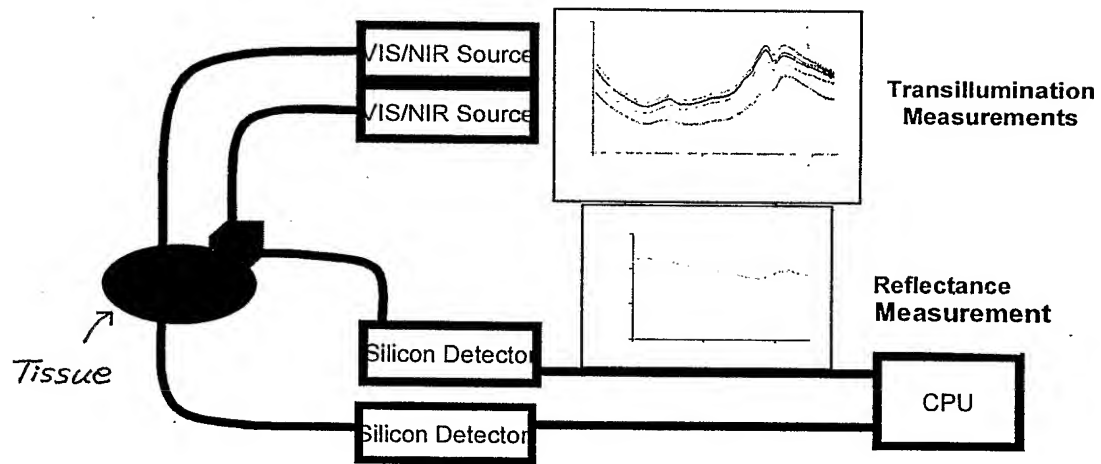


FIGURE 14

15/15

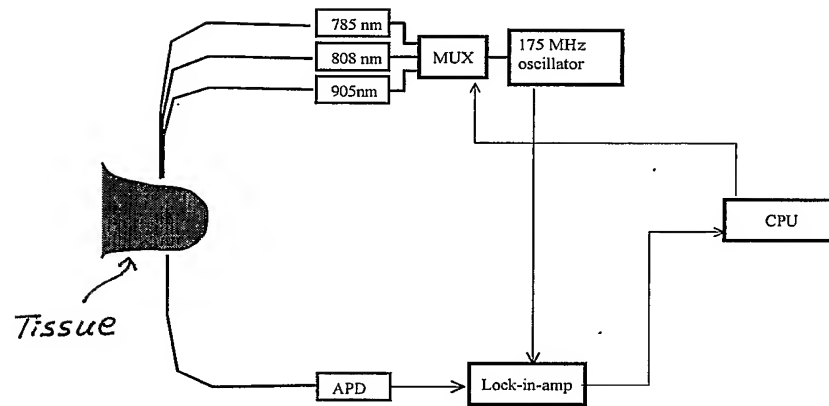


FIGURE 15

## INTERNATIONAL SEARCH REPORT

national Application No

PCT/CA 02/01771

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 A61B5/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 A61B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6 095 982 A (MITCHELL MICHELE FOLLEN ET AL) 1 August 2000 (2000-08-01) cited in the application column 7, line 10 -column 15, line 35; claims 12,13	16,18, 19,24
Y	EGAN AND DOLAN: "Optical Spectroscopy. Pre-mammography marker" ACTA RADIOLOGICA, vol. 29, no. 5, 1988, pages 497-503, XP008015249 cited in the application page 498, paragraph 4 -page 501, last paragraph	16,18, 19,24

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

## \* Special categories of cited documents :

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*&\* document member of the same patent family

Date of the actual completion of the international search

20 March 2003

Date of mailing of the international search report

28/03/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Manschot, J

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17

Claim 17 relates to an apparatus, but is only defined by method steps. It is not clear (see Article 6) how these method steps may characterize the apparatus and hence a meaningful search is not possible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CA 02/01771

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 1-15  
because they relate to subject matter not required to be searched by this Authority, namely:  
Rule 39.1(iv) PCT - Diagnostic method practised on the human or animal body
2. ☒ Claims Nos.: 17  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

national Application No

PCT/CA 02/01771

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6095982	A	01-08-2000	US 5697373 A	16-12-1997
			CA 2190374 A1	19-09-1996
			EP 0765134 A1	02-04-1997
			JP 10505167 T	19-05-1998
			WO 9628084 A1	19-09-1996
			US 5991653 A	23-11-1999
<hr/>				